FORUM STATISTICUM SLOVACUM Vol XVII (2021) No 2 pp 1-11

Rozcestník moderních přístupů směřujících k automatizované detekci anomálií v časových řadách
A guidepost to modern approaches directed towards automated anomaly detection in time series

Jitka Bartošová, Vladislav Bína, Vladimír Přibyl
Vysoká škola ekonomická v Praze, Fakulta managementu, Jarošovská 1117/II, 377 01 Jindřichův Hradec, Česká republika
University of Economics in Prague, Faculty of Management, Jarošovská 1117/II, 377 01 Jindřichův Hradec, Czech Republic
jitka.bartosova@vse.cz, vladislav.bina@vse.cz, vladimir.pribyl@vse.cz

Abstrakt: *S ohledem k velkému množství dat, která jsou k dispozici vedoucím pracovníkům v takřka každé současné organizaci se automatizované monitorování základních procesů zajišťujících chod organizací postupně stává naprostou nutností. Konečným posuzovatelem a rozhodovatelem musí vždy být manažer s příslušnou zodpovědností, nicméně alespoň pro základní a hrubou přípravu sestav informací o odchylkách (anomáliích, či alertech) v klíčových procesech lze využít širokou škálu matematických a statistických přístupů implementovaných ve vhodném prostředí pro zpracování dat, či ideálně přímo integrovaných v rámci informačního systému. Tento článek přináší alespoň základní přehled přístupů nejčastěji využívaných pro detekci anomálií.*

Abstract: *Given a large amount of data available to executives in almost every contemporary organization, automated monitoring of the basic processes in organizations is gradually becoming an absolute necessity. The final assessor and decision-maker must always be a manager with appropriate responsibilities. However, at least for basic and rough preparation of reports concerning deviations (anomalies or alerts) in key processes, we can use a wide range of mathematical and statistical approaches implemented in an appropriate data processing environment, or ideally integrated directly in the information system. This article provides at least a basic overview of the approaches frequently used for anomaly detection.*

Klíčová slova: *detekce anomálií, časové řady, rozklad, umělá inteligence*

Key words: *anomaly detection, time series, decomposition, artificial intelligence*

1 Úvod

Obrovský rozvoj informačních a komunikačních technologií v prvních dekádách 21. století otevřel podnikatelské a organizační sféře pestrou škálu nových přístupů a výzev. Poslední roky přinesly nástup moderních a slibných technologicky náročných možností v rámci širokých oblastí Průmyslu 4.0 a fenoménu big data. Patří sem přístupy, jakými jsou např. digitální dvojče, průmyslový internet věcí, adaptivní automatizace, robotizace, 3D tisk, chytrá logistika a chytrá výroba (viz např. Evjemo et al., 2020), ale i v době světové pandemie významně posilující oblast e-businessu (Elhrim et al., 2020). Některé z hojně skloňovaných moderních přístupů a zdánlivě slibných oblastí jsou dost

možná prázdnými pojmy a v testu času se neprosadí. Nicméně společným jmenovatelem je stále sílící potřeba zpracování velkého objemu dat, jež se neobejde bez alespoň částečného automatického předzpracování, které zefektivní či přímo umožní rozhodování příslušným manažerům.

Přirozeně s tím souvisí i rozvoj služeb nabízejících zpracování velkých objemů dat na částečně automatizované bázi. Vzhledem k tomu, že v podnikové a organizační sféře se z velké většiny jedná o činnost pravidelnou, nabývá na důležitosti sledování prosté skutečnosti, zda je chod jednotlivých procesů v organizaci v pořádku, či zda nedochází k nějakým zásadním výkyvům představujícím riziko pro fungování organizace. V případě výkyvů, či anomálií je pak klíčové dobré vymezení, hledání souvislostí a případně i kauzálních vztahů.

2 Automatizace hledání anomálií

Ukazuje se, že detekce odlehlých hodnot a anomálií v datech je komplexní a netriviální problém jak z hlediska lidského rozhodovatele, tak z hlediska počítačové podpory manažerského rozhodování (Tabesh et al., 2019). Pro člověka je zpravidla významnou potíží velké množství dat, na druhou stranu ale zpravidla se snadno řekne, která hodnota se jeví „normální“ a která je zřejmou „anomálií“. Do jisté míry komplementární je algoritmický přístup, velké množství dat není potíží, ale spolehlivost detekce a interpretace anomálních hodnot leckdy pokulhávají. Automatizovaná detekce anomálií je oblastí řešenou již několik desetiletí a mezi první dobře využitelné metody týkající se časových řad patřily např. Holt-Wintersovo exponenciální vyrovnávání (Chatfield, 1978) a ARIMA přístupy (Box a Jenkins, 1970).

Narůstající poptávka po řešení z oblasti detekce anomálií způsobila výrazný rozvoj nejen v oblasti teoretické, ale i v oblasti rozvoje komponent BI (business intelligence) řešení. Vlastní přístupy detekce anomálií přinesly velké společnosti jako je Twitter (Kejariwal, 2015), LinkedIn (Luminol, 2018), Facebook (Taylor a Letham, 2018), Google (Brodersen et al., 2015). Zpravidla byly systémy těchto společností volně k dispozici alespoň v nějaké průběžné variantě, mezi slibnými a velmi využívanými zůstává např. Prophet. Kromě toho na automatizované detekci anomálií alespoň částečně postavila svůj byznys řada analytických a BI společností, sem patří například společnost Anodot (Toledano et al., 2018, a https://www.anodot.com), nebo část BI řešení od firmy Microsoft: Microsoft Cognitive Service založená na Spectral Residual (SR) a Convolutional Neural Network (CNN) (viz Ren et al., 2019 a https://azure.microsoft.com/en-us/services/cognitive-services), či např. produkty společností Avora (https://avora.com) a CrunchMetrics (https://www.crunchmetrics.ai).

Ukazuje se, že zdůraznění, případně analýza neobvyklých a zajímavých událostí, které v určitém smyslu vybočují z dosavadního (či očekávaného) průběhu časové řady údajů, se stává velmi důležitou činností ve všech oblastech, kde je zpracováváno velké množství dat. Analýza se napříč oblastmi nasazení soustředí zejména na metriky podporující řízení důležitých podnikových procesů, zákaznických vztahů, či finančních ukazatelů a z toho plynou i vlastnosti časových řad, které obvykle agregují údaje na denní bázi, případně řídčeji a často vykazují týdenní a obvykle i další (např. roční) sezonní složky.

Metody detekce mohou jednak zkoumat časové řady jednotlivě a jednak se mohou zabývat skupinou časových řad jako celkem. V této části se budeme věnovat pouze řadám zahrnujícím jedinou proměnnou, struktury více časových řad budou zmíněny v kapitole 6. Přístupy k detekci anomálií lze dále členit na metody zohledňující časovou informaci a metody, které primárně časovou informaci nezahrnují. Dalším atributem používaných metod je i možnost použití detekce anomálií v reálném čase přímo při zpracování transakcí, což u drtivé většiny metod zahrnuje nutnost nového přepočítání modelu dané veličiny. Zároveň se vlastně jedná o výpočet založený na predikci budoucí hodnoty časové řady, na rozdíl od přístupů, které na základě modelu časové řady identifikují anomálie porovnáním celého reálného průběhu řady s jejím teoretickým modelem.

Při hledání anomálií v jednotlivých časových řadách tak můžeme využívat modelových přístupů (s modelováním hodnot v průběhu řady, či s predikcí budoucí hodnoty) a dále metody založené na hustotě jednotlivých hodnot. Do první kategorie spadají metody založené na regresním modelování, na dekompozici (např. STL, ETS) a také ARIMA přístupy, do druhé kategorie pak např. přístupy izolačních lesů, či přístupy založené na neuronových sítích.

3 Statistické přístupy k detekci anomálií

Chování časových řad zahrnuje řadu typických vzorů a základní přístupy k jejich modelování jsou založeny na rozdělení řad do základních komponent, kterými jsou trend, sezónní složka a cyklická složka. Při rozkladu časové řady se obvykle trendová a cyklická složka spojují do jediné „trend-cyklové“ složky, která se zpravidla jednoduše (a mírně nepřesně) označuje jako trend. Obvyklým způsobem tedy je rozklad na trend (myšleno trend-cyklus), sezónní složku a reziduální složku (tedy část variability dat, kterou model nedokáže vysvětlit). V případě řad s větší frekvencí (nejméně denní), může popis zahrnovat i více sezónních komponent odpovídajících různé periodicitě.

3.1 ETS dekompozice

ETS rozklad vychází z přístupů vytvořených ke konci padesátých let a založených na exponenciálním vyhlazování s využitím vážených průměrů předchozích pozorování. Samotné exponenciální vyrovnávání (Brown, 1959) bylo rozšířeno v přístup zahrnující modelování lineárního trendu (Holt, 1957), který byl dále rozšířen o modelování sezónní složky (Winters, 1960). Postupem času byla rozvinuta široká škála ETS modelů, jejich přehled lze nalézt např. u Hyndmana a Khandakara (2008). K identifikaci anomálií se využívá namodelované hodnoty časové řady a predikčních intervalů s uživatelsky zvolenou hladinou spolehlivosti.

3.2 STL dekompozice

Jedním z univerzálních a poměrně robustních přístupů je dekompozice STL (Cleveland a Cleveland, 1990), která je založena na klasickém rozkladu na sezónní a trendovou složku, přičemž trendová část je modelována s využitím LOESS odhadu (Cleveland, 1979). Na rozdíl od některých dalších přístupů (jako je X11, či SEATS dekompozice), tento rozklad umí pracovat s obecnějšími typy sezónních komponent (ne pouze s čtvrtletními nebo měsíčními) a je poměrně robustní s ohledem k odlehlým hodnotám, tedy jednotlivá pozorování nemají výrazný vliv na odhady složek (Hyndman a Athanasopoulos, 2018). Zároveň se sezónní složka může v čase měnit a rychlost změny spolu s hladkostí trendové složky je možné v některých implementacích měnit. Díky využití LOESS odhadu je metoda velmi úspěšná v popisu trend-cyklové složky, tedy ve chvílích, kdy je zřetelný dlouhodobý trend. Z uvedených důvodů tato a následující metoda patří mezi základní používané přístupy pro detekci anomálií.

3.3 Twitter dekompozice

Jedná se o přístup velmi podobný STL rozkladu, využívá stejný přístup k práci se sezónní složkou a je použit v balíčku AnomalyDetection společnosti Twitter (Hochenbaum et al., 2017). Na rozdíl od STL však trendovou složku modeluje s využitím (lokálního výpočtu) mediánu. Z toho důvodu funguje velmi dobře, pokud je dlouhodobý trend méně významný a dominují krátkodobé sezónní složky.

Dekompozice STL a Twitter lze použít se dvěma přístupy k identifikaci anomálních bodů. A sice s metodami: IQR (Inner Quartile Range) a GESD (Generalized Extreme Studentized Deviate test). Metoda IQR využívá mezikvartilové rozpětí a v základním nastavení jsou anomálie identifikovány za hranicí vzdálenosti trojnásobku od mezikvartilového rozpětí. Jedná se o rychlý přístup, který nevyužívá iterativního výpočtu, jeho slabou stránkou je horší adaptabilita. Flexibilnější GESD přístup dokáže dobře pracovat i s významnými

anomáliemi, které v průběhu výpočtu iterativně odstraňuje. Proto je výpočetně náročnější než IQR přístup, což je patrné zejména v případě většího množství časových řad s velkým množstvím pozorování.

3.4 ARIMA modely

Spolu s přístupem exponenciálního vyrovnávání patří mezi nejpoužívanější metody modelování časových řad. Na rozdíl od dekompozičních přístupů však ARIMA metody nevyužívají popisu trendu a sezónnosti, ale autokorelačních charakteristik modelovaných dat.

Modely Autoregressive Integrated Moving Average zahrnují jednak autoregresní charakter modelu (tedy závislost aktuálně pozorované hodnoty na určeném počtu starších hodnot, jednak integrovanou část spočívající ve využití diferencí tak, aby charakter řady byl stacionární a složku klouzavých průměrů sloužící k modelování aktuálních hodnot jako lineární kombinace chyb několika předchozích hodnot. Přístupy sezonních modelů zahrnují uvedené složky i pro sezónní komponentu časové řady.

Vzhledem k tomu, že plánovaný účel nasazení vyžaduje co nejvyšší míru automatizace, využíváme přístup pojmenovaný jako „automatizovaná ARIMA“ (Hyndman a Khandahar, 2008). Tento přístup využívá testování jednotkového kořene společně s minimalizací AIC a MLE kritérií pro výběr nejvhodnějšího ARIMA modelu. Opět se jedná o univerzální a široce využitelný přístup.

3.5 Prophet

Facebookovský Prophet (Taylor a Letham, 2018) patří rovněž mezi aditivní dekompoziční přístupy. Využívá po částech lineární křivky a logistické přechody k modelování neperiodické složky, dále model zahrnuje sezónní složku a složku zahrnující uživatelem vložené informace o svátcích způsobujících odchylky v časové řadě. Neperiodická složka je odhadována s využitím regresních principů, složka sezónní je modelována s využitím Holt-Wintersova exponenciálního vyrovnávání (Chatfield, 1978). Protože se jedná o přístup k modelování a predikci časových řad, je jeho využití pro detekci anomálií přímočaré. Prophet patří mezi velmi nadějné přístupy, protože se jedná o velmi robustní přístup dobře zvládající chybějící hodnoty, velké sezónní změny, výkyvy v době svátků a prázdnin i výrazné posuny v trendu.

3.6 TBATS modely

Výše uvedené dekompoziční modely v zásadě uvažovaly pouze jednoduchý typ sezónní složky. Mezi přístupy umožňující zahrnout vliv více sezón najednou (v našem případě zpravidla týdenní a roční) patří TBATS modely (De Livera et al., 2011). Pro výpočet používají kombinaci modelu exponenciálního vyhlazování spolu s využitím fourierovských členů k modelování periodických složek a Box-

Coxovy transformace. Na rozdíl od přístupů harmonické regrese TBATS modely umožňují mírnou změnu sezonních komponent s vývojem řady v čase. Tato komplexnost a flexibilita modelů může být velmi užitečná, nicméně využitelnost je výrazně omezena faktem, že v případě delších časových řad je tvorba modelu velmi výpočetně náročná.

4 Metody z oblasti umělé inteligence

Pro detekci anomálií jsou využívány rovněž přístupy zložené na učení bez učitele, jakými jsou například moderní a často používané izolační lesy, či specializované přístupy v rámci umělých neuronových sítí (Recurrent neural networks a vylepšené deep-learningové LSTM autoencoders), nebo googlovský přístup CausalImpact, či využití konvolučních neuronových sítí v řešení společnosti Microsoft. Praktickým úskalím těchto metod může být to, že není k dispozici jejich otevřená implementace vhodná pro zapojení do vyvíjeného systému a nemusí být kompatibilní se zvolenou softwarovou architekturou.

4.1 Izolační lesy

Izolační lesy (Liu et al., 2008) představují přístup učení bez učitele obecně aplikovaný v oblasti detekce anomálií. Princip metody je založen na sestavení souboru izolačních stromů a přes tento soubor je pak průměrováno vypočtené anomální skóre jednotlivých pozorování charakterizující délku cesty k tomuto pozorování. Základní myšlenkou je to, že “normální” body jsou svou hodnotou obdobné jako řada dalších pozorování, tedy mají v sousedství řadu dalších hodnot. K rozlišení takových bodů bude potřeba v drtivé většině vygenerovaných stromů projít velkým množstvím větvení. Naproti tomu se v případě odlehlých hodnot podaří je separovat od ostatních zpravidla průchodem jen několika málo větví. Průměrovaná hloubka průchodu k dané hodnotě stromem tak po použití vhodné normalizace umožňuje identifikovat odlehlé hodnoty.

V případě, že se jedná o identifikaci anomálie v časových řadách, obvykle je potřeba použít vhodné rozšíření této metody tak, aby zohledňovala časovost dat. Jednou z možností je využití klouzavého okna (Ding a Fei, 2013) k zohlednění lokálnosti, resp. časové souvislosti. Nejjednodušší variantou přitom je využití jedné, či několika zpožděných časových řad jako dalších dimenzí pro detekci s využitím základního algoritmu.

4.2 LSTM autoencodery

Neuronové sítě zvané autoencodery reprezentují další přístup učení bez učitele a jsou využívané jako efektivní metody pro kompresi dat. Kromě toho jsou užitečné jako prostředky pro redukci dimenzionality, zobecňování a zpracování obrazů. Detekce anomálií s využitím autoencoderů je založená na tom, že natrénovaný autoencoder dokáže běžná data dobře rekonstruovat, naproti

tomu fakt, že autoencoder v rekonstrukci neuspěje, znamená, že se velmi pravděpodobně jedná o anomální data.

V případě časových řad (a jiných sekvenčních dat) lze využívat LSTM (Long Short-Term Memory) autoencodery (Gers et al., 1999, případně Kulanuwat et al., 2021), které nejsou zatížené nedostatky rekurentních neuronových sítí týkajících se selhání učících algoritmů na mizejících či naopak rostoucích gradientech dlouhých časových řad.

4.3 CausalImpact

CausalImpact (Brodersen et al., 2015) je přístupem využívaným společností Google a je založen na ekonometrických principech Bayesovských strukturálních rovnic pro časové řady. Přístup s využitím časových řad z kontrolní skupiny vytváří syntetickou základní časovou řadu. Anomálie pak nacházíme porovnáním nové, tzv. testové řady s takto vytvořenou syntetickou základní řadou.

5 Moderní implementace důležitých přístupů

V rámci předchozí kapitoly jsme předložili přehled přístupů pro detekci anomálií časových řad a vytipovali sadu dostatečně pokročilých (a tedy i nadějných) přístupů. Alespoň základní možnosti implementace těchto metod v prostředí statistického softwaru R (R Core Team, 2021) se zaměřením na ETS, Arima, STL, Twitter a IF, částečně i na Prophet a TBATS přístupy přinášíme zde.

5.1 Knihovna forecast

Jako základní východisko pro řadu modelů lze využít knihovnu **forecast** (Hyndman et al., 2018). Tato knihovna zahrnuje mimo jiné moderní a robustní implementace přístupů ARIMA (včetně automatizované tvorby modelů přístupem auto.arima), ETS přístupy, STL a Twitter dekompozice i implementaci TBATS modelů a umožňuje na základě vytvořených modelů jednorozměrných časových řad generovat predikční intervaly namodelovaných hodnot, které slouží k identifikaci anomálií.

5.2 Knihovna anomalize

Pro implementaci STL a Twitter přístupu je dobře použitelná knihovna **anomalize** (Dancho a Vaughan, 2020), která umožňuje využívat IQR a GESD přístupy pro detekci anomálních bodů i další parametrizaci.

5.3 Knihovna prophet

Pro využití robustního facebookovského Prophet přístupu dobře zvládajícího sezonní výkyvy, zahrnutí svátků a změny v trendech slouží v prostředí R otevřená knihovna **prophet** (Taylor a Letham, 2020).

5.4 Knihovna solitude jako implementace izolačních lesů

Jako implementace přístupu izolačních lesů může sloužit R balíček **solitude** (Liu et al., 2012). Jedná se o moderní implementaci dobře kompatibilní s pokročilými datovými strukturami v R a využívající paralelizace výpočtu souboru izolačních stromů. Významným úskalím tohoto přístupu je nekorektní zacházení s časovou dimenzí, které lze částečně překonat využitím zpožděné časové řady jako další dimenze. Dalším úskalím je fakt, že není dobře standardizované hraniční skóre oddělující anomální hodnoty. Před případným využitím této metody v oblasti detekce anomálií je tak nezbytné navrhnout vhodnou standardizaci.

5.5 Moderní práce s časovými údaji: lubridate a tsibble

Flexibilní a korektní zpracování různých formátů času a data včetně převodů mezi textovými a datovými formáty, extrakce a nastavování komponent, aritmetických operací a zaokrouhlování, ale i práci s časovými zónami, přestupnými roky a letním časem, to vše umožňuje moderní knihovna **lubridate** (Grolemund a Wickham, 2011).

Moderním přístupem k manipulaci s daty v prostředí R je sbírka balíků **tidyverse** (Wickham et al., 2019, nebo Wickham a Grolemund, 2016). Umožňuje import dat, manipulaci, čištění a úpravu, vizualizaci s vhodným a cíleným využíváním doménově specifických balíčků v této sbírce. Sbírku **tidyverse** rozšiřuje a efektivní práci s časovými daty umožňuje balíček **tsibble** (Wang et al., 2020). Rozšiřuje vlastnosti obecných **tibble** datových struktur, přičemž užívá časové údaje jako klíčové datové sloupce a využívá pořadí daného plynutím času.

6 Základní možnosti výběru významných anomálií a redukce dimenzionality

Výše popsané automatizované přístupy detekce anomálií společně s nasazením na velké množství dat vznikajících a archivovaných v běžné praxi větších firem a organizací představují zdroj detekce velkého množství výkyvů a anomálií, na které může být uživatel systému upozorněn. To samozřejmě přináší velkou pracovní zátěž při kontrole výstupů, či naopak riziko přehlédnutí některých významných alertů mezi spoustou upozornění podstatně méně podstatných. Pro zpřehlednění, vyšší uživatelskou přívětivost a lepší využitelnost systému lze uvažovat následující tři základní přístupy.

6.1 Ekonomické souvislosti

Východiskem pro seskupování na základě ekonomických ukazatelů je řada ekonomických vztahů a souvztažností od jednoduchých po komplexnější a zároveň od deterministických až po stochastické. Velkým úskalím tohoto přístupu je fakt, že jeho důslednější aplikace zpravidla narazí na specifičnost

sestav ukazatelů v různých oblastech a oborech a bude tudíž vyžadovat implementaci diferencovanou pro různá nasazení alertovacího systému. Nicméně typicky nejzákladnější ekonomické souvislosti typu výkyvů ve výši tržeb a s tím zároveň v zásadě zbytečně reportovaným výkyvům v odvedené DPH lze snadno předcházet.

6.2 Metody hledající příbuzné anomálie na základě statistických souvislostí

Statistické přístupy umožňují průběžně vyhledávat úzce související řady, které nebyly zjištěny pomocí známých ekonomických souvislostí. Výhodou přístupu je to, že může být do značné míry automatizován. Časové řady mohou být zjišťovány pomocí korelačních a kointegračních přístupů a případně i pokročilejšími ekonometrickými přístupy, jakými je např. Grangerova analýza (Granger, 1969), nebo Convergent cross mapping (Čenys et al., 1991).

6.3 Hierarchické modely časových řad a sbalování souvisejících anomálií

Ve většině větších společností existuje struktura středisek a poboček a např. z hlediska vyráběného, či prodávaného sortimentu lze velmi často identifikovat zpravidla vícestupňovou hierarchickou strukturu výrobků. Jednoduchou hierarchii nebo i vícenásobné seskupení časových řad lze modelovat přístupy uvedenými v práci Wickramasuriya et al. (2019). Velkou výhodou těchto přístupů je to, že modely těchto časových řad jsou spolu v souladu, to znamená, že např. i součet predikovaných hodnot za nižší jednotky plně odpovídá predikované hodnotě na vyšší úrovni. Při využití těchto přístupů může uživatel procházet hierarchií od vyšší úrovně po nižší a detekované anomálie postupně rozbaluje podle potřeby, přičemž na vyšší úrovni mohou být indikovány např. pouze výrazné anomálie, případně i s indikací, jak velké množství anomálií je na vnořených úrovních.

7 Závěr

Tento článek přináší základní shrnutí metod využitelných pro automatickou detekci anomálií a vytváření varování v rámci informačních systémů jako podkladu pro rozhodování vedoucích pracovníků. V návaznosti na rychlý rozvoj v oblasti dat a informací v podnikové sféře se zejména v posledním desetiletí překotně rozvíjejí i přístupy k predikci v časových řadách a hledání neočekávaných výkyvů v jejich průběhu. Kromě tradičních matematicko-statistických přístupů vznikají moderní přístupy šité na míru internetovému prostředí a prostředí sociálních sítí. Základní přehled uvedený v tomto článku může sloužit prvotní orientaci v této oblasti a lze jej využít v počátcích přípravy automatizovaného systému. Vždy je ale potřeba mít na paměti doménu využití, příslušné ekonomické závislosti, strukturu a hierarchii časových dat i jejich

periodicitu, sezónnost a cykličnost. Tomu je pak zapotřebí přizpůsobit výběr a propojení použitých metod.

8 Literatura

Box, G., Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Brodersen, K. H. et al. (2015). *Inferring causal impact using Bayesian structural time-series models.* Annals of Applied Statistics, 9(1), 247-274.

Brown, R. G. (1959). *Statistical forecasting for inventory control.* McGraw/Hill.

Chatfield, C. (1978). *The Holt-winters forecasting procedure.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 27(3), 264-279.

Cleveland, R. B. et al. (1990). *STL: A seasonal-trend decomposition.* Journal of Official Statistics, 6(1), 3-73.

Cleveland, W. S. (1979). *Robust locally weighted regression and smoothing scatterplots.* Journal of the American Statistical Association, 74(368), 829-836.

Čenys, A., Lasiene, G., Pyragas, K. (1991). *Estimation of interrelation between chaotic observables.* Physica D: Nonlinear Phenomena, 52(2-3), 332-337.

Dancho, M., Vaughan, D. (2020). *anomalize: Tidy Anomaly Detection. R package version 0.2.2.* https://CRAN.R-project.org/package=anomalize.

De Livera, A. M., Hyndman, R. J., Snyder, R. D. (2011). *Forecasting time series with complex seasonal patterns using exponential smoothing.* Journal of the American Statistical Association, 106(496), 1513-1527.

Ding, Z., Fei, M. (2013). *An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window.* IFAC Proceedings Volumes, 46(20), 12-17.

Elrhim, M. A., Elsayed, A. (2020). *The Effect of COVID-19 Spread on the e-commerce market: The case of the 5 largest e-commerce companies in the world.* https://ssrn.com/abstract=3621166.

Evjemo, L. D. et al. (2020). *Trends in Smart Manufacturing: Role of Humans and Industrial Robots in Smart Factories.* Current Robotics Reports, 1(2), 35-41.

Gers, F. A., Schmidhuber, J., Cummins, F. (1999). *Learning to forget: Continual prediction with LSTM.* In 9th International Conference on Artificial Neural Networks: ICANN '99 (pp. 850-855).

Granger, C. W. J. (1969). *Investigating causal relations by econometric models and cross-spectral methods.* Econometrica: Journal of the Econometric Society, 424-438.

Grolemund, G., Wickham, H. (2011). *Dates and Times Made Easy with lubridate.* Journal of Statistical Software, 40(3), 1-25. https://www.jstatsoft.org/v40/i03.

Hochenbaum, J., Vallis, O. S., Kejariwal, A. (2017). *Automatic anomaly detection in the cloud via statistical learning.* https://arxiv.org/abs/1704.07706.

Hyndman, R. J., Athanasopoulos, G. (2018). *Forecasting: principles and practice.* OTexts. https://otexts.com/fpp3.

Hyndman, R. J. et al. (2018). *forecast: Forecasting functions for time series and linear models.* https://pkg.robjhyndman.com/forecast.

Hyndman, R. J., Khandakar, Y. (2008). *Automatic time series forecasting: the forecast package for R.* Journal of Statistical Software, 27(3), 1-22.

Holt, C. C. (1957). *Forecasting Trends and Seasonals by Exponentially Weighted Averages*, vol. 52. Carnegie Institute of Technology, Pittsburgh, Pa, USA.

Kejariwal, A. (2015). *Introducing practical and robust anomaly detection in a time series. Twitter Engineering Blog.* https://blog.twitter.com/engineering/en_us/a/2015/introducing-practical-and-robust-anomaly-detection-in-a-time-series.html.

Kulanuwat, L. et al. (2021). *Anomaly detection using a sliding window technique and data imputation with machine learning for hydrological time series.* Water, 13(13), 1862.

Liu, F. T., Ting, K. M., Zhou, Z. H. (2012). *Isolation-based anomaly detection.* ACM Transactions on Knowledge Discovery from Data (TKDD), 6(1), 1-39.

Luminol (2018) *Luminol: LinkedIn's Anomaly Detection and Correlation Library.* https://github.com/linkedin/luminol.

R Core Team (2021). R: *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Ren, H. et al.(2019). *Time-series anomaly detection service at Microsoft.* In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 3009-3017).

Tabesh, P., Mousavidin, E., Hasani, S. (2019). *Implementing big data strategies: A managerial perspective.* Business Horizons, 62(3), 347-358.

Taylor, S. J., Letham, B. (2018). *Forecasting at scale.* The American Statistician, 72(1), 37-45.

Toledano, M. et al. (2018). *Real-time anomaly detection system for time series at scale.* In KDD 2017 Workshop on Anomaly Detection in Finance (pp. 56-65).

Wang, E., Cook, D., Hyndman, R. J. (2020). *A new tidy data structure to support exploration and modeling of temporal data.* Journal of Computational and Graphical Statistics, 29(3), 466-478.

Wickham, H. e al. (2019). *Welcome to the tidyverse.* Journal of Open Source Software, 4(43), 1686.

Wickham, H., Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data.* O'Reilly Media, Inc.

Wickramasuriya, S. L., Athanasopoulos, G., Hyndman, R. J. (2019). *Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization.* Journal of the American Statistical Association, 114(526), 804-819.

Winters, P. R. (1960). *Forecasting sales by exponentially weighted moving averages.* Management Science, 6(3), 324-342.

9 Poděkování

Tvorba tohoto článku byla podpořena Technologickou agenturou České republiky v programu TREND v rámci projektu FW01010606: Adaptivní alertovací business intelligence systém.